

Ajtai-Szemerédi Theorems over quasirandom groups

Tim Austin

Abstract Two versions of the Ajtai-Szemerédi Theorem are considered in the Cartesian square of a finite non-Abelian group G . In case G is sufficiently quasirandom, we obtain strong forms of both versions: if $E \subseteq G \times G$ is fairly dense, then E contains a large number of the desired patterns for most individual choices of ‘common difference’. For one of the versions, we also show that this set of good common differences is syndetic.

1 Introduction

A classical result of Ajtai and Szemerédi [1] asserts the following. For every $\delta > 0$ there is an $N_0 < \infty$ such that, if G is a finite cyclic group with $|G| \geq N_0$, and $E \subseteq G \times G$ satisfies

$$|E| \geq \delta |G|^2, \quad (1)$$

then there are $x, y \in G$ and $r \in G \setminus \{0\}$ such that

$$E \supseteq \{(x, y), (x + r, y), (x, y + r)\}. \quad (2)$$

A set satisfying (1) is called δ -**dense** in $G \times G$. Subsets of $G \times G$ of the kind on the right-hand side of (2) are called **Abelian corners**.

Several proofs of this are now known, and it has been generalized to subsets of $G \times G$ for arbitrary finite Abelian groups G : see, for instance, the discussions around [21, Propositions 10.47 and 11.28].

Tim Austin
 Courant Institute, New York University, New York, NY 10012, United States of America
 e-mail: tim@cims.nyu.edu

All known proofs actually give a stronger conclusion: that there is some $c > 0$ depending only on δ with the property that any δ -dense set $E \subseteq G \times G$ must satisfy

$$|\{(x, y, r) \in G^3 : \{(x, y), (x + r, y), (x, y + r)\} \subseteq E\}| \geq c|G|^3. \quad (3)$$

This paper considers two candidate generalizations of these results to non-Abelian groups G . The most obvious non-Abelian analogs of Abelian corners are triples of the form

$$\{(x, y), (gx, y), (x, gy)\}, \quad g \neq e, \quad (4)$$

where e is the identity of G . We refer to these as **naïve corners**. A second possibility is triples of the form

$$\{(x, y), (gx, y), (gx, gy)\}, \quad g \neq e. \quad (5)$$

Triples of this second kind first appeared in [5] (in the related setting of infinite amenable groups and their probability-preserving actions), so we refer to them as **Bergelson-McCutcheon-Zhang ('BMZ') corners**. Solymosi [20] has studied BMZ corners in finite groups in; he writes them as

$$\{(x, y), (xg, y), (x, gy)\}, \quad g \neq e,$$

but these may be identified with BMZ corners by first writing in terms of y, g and $x' := xg$ and then applying the transformation $(a, b) \mapsto (a^{-1}, b)$.

When G is Abelian, naïve corners are Abelian corners, and BMZ corners are equivalent to them by a simple change of variables, but this is not true for non-Abelian G . As discussed in [5], when searching for BMZ corners inside a δ -dense subset of $G \times G$, some methods from the Abelian setting have useful generalizations which do not seem to apply to naïve corners.

Either of these patterns, and indeed many others, must appear inside a δ -dense subset of $G \times G$ for any sufficiently large finite group G . This is simply because such a G must contain a fairly large Abelian subgroup [11], to whose cosets one can apply the known results for Abelian groups.

Our first theorem states this result formally. Here and for the rest of the paper, we also allow G to be a non-finite compact metrizable group, since our methods handle these with little change. We generally refer to these simply as ‘compact’ groups, suppressing the assumption that they are metrizable. If G is a compact group then its Haar probability measure is denoted by m_G ; if $H \leq G$ is a closed subgroup and $g \in G$, then m_{gH} and m_{Hg} denote respectively the left- and right-translates of m_H , regarded as a measure on G , by the element g .

Theorem A. *For any $\delta > 0$ and any $k \in \mathbb{N}$ there is an $N_0 < \infty$ for which the following holds. If G is a compact group with $|G| \geq N_0$ (in particular, if G is not finite), and $E \subseteq G^k$ is Borel and satisfies $m_{G^k}(E) \geq \delta$, then there are*

$$(x_1, \dots, x_k) \in G^k \quad \text{and} \quad g \in G \setminus \{e\}$$

such that

$$E \supseteq \{(x_1^{\eta_1}, \dots, x_k^{\eta_k}) : (\eta_1, \dots, \eta_k) \in \{0, 1\}^k\},$$

where for each $i \leq k$ we set $x_i^0 := x_i$ and $x_i^1 := gx_i$.

The analogous result in which we set $x_i^1 := x_i g$ follows immediately, simply by applying Theorem A to the set

$$\{(y_1^{-1}, \dots, y_k^{-1}) : (y_1, \dots, y_k) \in E\}.$$

The simple proof of Theorem A will be given in Section 2. However, the proof does not give a generalization of the ‘counting’ result of inequality (3) to non-Abelian G , because it counts only those patterns for which the parameter g lies in some Abelian subgroup of G , which may be very small compared to G itself.

It seems important to begin this paper with Theorem A, but our main results are of a different kind. These assert that, if G is sufficiently quasirandom (as defined in [13]) in terms of δ , and $E \subseteq G \times G$ is δ -dense, then there is a large set of possible ‘common differences’ g such that E must contain both many naïve corners and many BMZ corners with that specific choice of g . In contrast to the Abelian setting, the assumption of quasirandomness allows one to obtain reasonable and explicit bounds. The current best bounds in the Abelian case are those of Shkredov [18, 19].

Theorem B. *Let $\varepsilon > 0$ and suppose that G is a D -quasirandom compact group. Let $E \subseteq G \times G$ be Borel and set $\delta := m_{G \times G}(E)$. Then the set*

$$A := \{g \in G : m_{G \times G}\{(x, y) : \{(x, y), (gx, y), (x, gy)\} \subseteq E\} \geq \delta^3 - \varepsilon\}$$

has $m_G(A) \geq 1 - 2\sqrt{3}D^{-1/4}/\varepsilon$. In particular, if $D \gg (1/\varepsilon)^4$ then A is most of G .

This is proved in Section 4. As far as I know, this is the first result for naïve corners in non-Abelian groups, although the proof uses only some simple estimates and classical representation theory.

Similar methods give an analogous result for BMZ corners, although the bounds we obtain are more complicated, and it seems unlikely that they are optimal. In formulating the results for BMZ corners, we no longer take care over the exact values of constants: universal constants are simply denoted by $O(1)$.

Theorem C. *For any $\varepsilon \in (0, 1/2)$ there is a*

$$D = \exp((1/\varepsilon)^{O(1)}) \tag{6}$$

for which the following holds. Let G be a D -quasirandom compact group, let $E \subseteq G \times G$ be Borel, and set $\delta := m_{G \times G}(E)$. Then the set

$$B := \{g \in G : m_{G \times G}\{(x, y) : \{(x, y), (gx, y), (gx, gy)\} \subseteq E\} \geq \delta^4 - \varepsilon\}$$

has $m_G(B) \geq 1 - \varepsilon$.

In the setting of Theorem C, a closely related argument proves another fact about the set B of ‘good common differences’.

Theorem D. *For any $\varepsilon \in (0, 1/2)$ there is a*

$$D = \exp \left(\exp \left((1/\varepsilon)^{O(1)} \right) \right) \quad (7)$$

for which the following holds. If G is a D -quasirandom compact group and $E \subseteq G \times G$ is Borel, then the set B defined in Theorem C is K -left-syndetic for some

$$K = \exp \left((1/\varepsilon)^{O(1)} \right).$$

This gives a different sense in which the set B is ‘large’. We henceforth abbreviate ‘ K -left-syndetic’ to ‘ K -syndetic’.

Qualitative versions of Theorems C and D have recently been proved by Bergelson, Robertson and Zorin-Kranich [6]. Their work uses ergodic theory for actions of certain ultraproducts of sequences of finite groups, which are of course highly infinite; it gives no explicit control on D or K in terms of ε . Our proof below uses only elementary inequalities and representation theory, so should be simpler for the reader not versed in non-standard analysis. In this respect, the present paper relates to [6] as did the earlier work [3] to [7]. We prove Theorems C and D in Section 5.

Theorems B, C and D together raise the following.

Question 1. Is it true that, if G is sufficiently quasirandom in terms of $\varepsilon \in (0, 1/2)$, then the set A from Theorem B is K -syndetic for some K depending only on ε ? \triangleleft

Intuitively, quasirandom groups are very far from Abelian groups. Taken together, the Ajtai-Szemerédi Theorem for Abelian groups and our Theorems B and C make it natural to ask whether a lower bound of the kind in (3) holds for either naïve or BMZ corners and for all finite groups G that are sufficiently large.

In fact, this is true for BMZ corners, since the original graph-theoretic proof of Ajtai and Szemerédi generalizes fairly easily to those: see [20]. This observation has a counterpart in ergodic theory, where related questions about multiple recurrence have been studied [5, 4, 10]. This is why the pattern (5) was first introduced into ergodic theory by Bergelson, McCutcheon and Zhang in [5].

However, I do not see a way to apply that graph-theoretic argument in looking for naïve corners, so for those the following question may be of interest.

Question 2. Is it true that for every $\delta > 0$ there is a $c > 0$ such that, whenever G is a compact group and $E \subseteq G \times G$ satisfies $m_{G^2}(E) \geq \delta$, one also has

$$m_{G^3} \{ (x, y, g) \in G^3 : \{ (x, y), (gx, y), (x, gy) \} \subseteq E \} \geq c?$$

\triangleleft

2 Proof of Theorem A

Proof (Theorem A). Step 1: finite groups. The theorem is well-known among finite Abelian groups: see, for instance, [21, Proposition 11.28]. Fix $\delta > 0$ and $k \in \mathbb{N}$, and let n be so large that, if A is any finite Abelian group of order at least n and $F \subseteq A^k$ satisfies $m_{A^k}(F) \geq \delta$, then F must contain a pattern of the kind in Theorem A.

By a classical result of Erdős and Straus [11], we may now choose N so large that any finite group G of order at least N contains an Abelian subgroup H of order at least n (see also [15] for an essentially optimal estimate of N in terms of n). Suppose that $E \subseteq G^k$ has $m_{G^k}(E) \geq \delta$. Then

$$m_{G^k}(E) = \int \cdots \int m_{Hg_1 \times \cdots \times Hg_k}(E) m_{H \setminus G}(\mathrm{d}(Hg_k)) \cdots m_{H \setminus G}(\mathrm{d}(Hg_1)),$$

so there are some cosets Hg_1, \dots, Hg_k for which

$$m_{Hg_1 \times Hg_2 \times \cdots \times Hg_k}(E) \geq \delta.$$

Define $F \subseteq H^k$ by

$$E \cap (Hg_1 \times Hg_2 \times \cdots \times Hg_k) = F \cdot (g_1, \dots, g_k).$$

Applying the Abelian case of the theorem to F gives some

$$(x_1, \dots, x_k) \in H^k \quad \text{and} \quad h \in H \setminus \{e\}$$

such that

$$F \supseteq \{(x_1^{\eta_1}, \dots, x_k^{\eta_k}) : (\eta_1, \dots, \eta_k) \in \{0, 1\}^k\},$$

with the obvious analog of the notation in the statement of Theorem A. Translating back to E , this gives

$$E \supseteq \{((x_1 g_1)^{\eta_1}, \dots, (x_k g_k)^{\eta_k}) : (\eta_1, \dots, \eta_k) \in \{0, 1\}^k\}.$$

Step 2: other compact Lie groups. If G is a non-finite compact Lie group, then it contains a nontrivial toral subgroup. This, in turn, contains finite cyclic subgroups of arbitrarily large cardinality. Letting H such a subgroup of cardinality at least n , we may complete the proof as in Step 1.

Step 3: general compact groups. Finally, let G be an arbitrary compact group which is not finite or a Lie group. As a standard consequence of the Peter-Weyl Theorem [8, Section III.3], there are a continuous surjective homomorphism $\pi : G \longrightarrow \overline{G}$ to a compact Lie group such that the set

$$\overline{E} := \{(\overline{x}_1, \dots, \overline{x}_k) \in \overline{G}^k : m_{\pi^{-1}\{\overline{x}_1\} \times \cdots \times \pi^{-1}\{\overline{x}_k\}}(E) > 1 - 2^{-k-1}\} \quad (8)$$

satisfies

$$m_{G^k}(E \triangle (\pi \times \cdots \times \pi)^{-1}(\overline{E})) < \delta/2,$$

and hence $m_{\overline{G}^k}(\overline{E}) > \delta/2$.

Since G is not finite or a Lie group, we may choose \overline{G} to have cardinality as large as we please (allowing infinity if necessary). Having done so, either Step 1 or Step 2 gives some

$$(\overline{x}_1, \dots, \overline{x}_k) \in \overline{G}^k \quad \text{and} \quad \overline{g} \in \overline{G} \setminus \{\overline{e}\}$$

such that

$$\overline{E} \supseteq \{(\overline{x}_1^{\eta_1}, \dots, \overline{x}_k^{\eta_k}) : (\eta_1, \dots, \eta_k) \in \{0, 1\}^k\},$$

with the obvious analog of the notation in the statement of Theorem A.

Finally, consider lifts x_1, \dots, x_k and g chosen independently at random from the Haar measures $m_{\pi^{-1}\{\overline{x}_1\}}, \dots, m_{\pi^{-1}\{\overline{x}_k\}}$ and $m_{\pi^{-1}\{\overline{g}\}}$. Observe that each gx_i is then a random lift of \overline{gx}_i with distribution $m_{\pi^{-1}\{\overline{gx}_i\}}$. Define each $x_i^{\eta_i}$ using these lifts in the usual way. Then it follows from the definition (8) that each of the events

$$(x_1^{\eta_1}, \dots, x_k^{\eta_k}) \in E \quad \text{for } (\eta_1, \dots, \eta_k) \in \{0, 1\}^k$$

has probability at least $1 - 2^{-k-1}$. Therefore, by the first-moment bound, there is some choice of x_1, \dots, x_k and g for which all of these events occur simultaneously. This completes the proof. \square

3 Preliminary discussion of the results for quasirandom groups

The proofs of Theorems B, C and D have several common elements. This section introduces some of those.

It will be convenient to use some simple notation and terminology from ergodic theory. Given a compact group G , a **probability** $(G \times G)$ -**space** is a tuple (Y, Σ, ν, S, T) in which (Y, Σ, ν) is a probability space and S and T are two commuting, ν -preserving G -actions on that space. Since they commute, S and T together define an action of $G \times G$, hence the name. Given such a probability $(G \times G)$ -space, we will write Σ^S , Σ^T and $\Sigma^{S,T}$ for the σ -subalgebras of sets in Σ that are invariant under S , T , or the whole $(G \times G)$ -action, respectively.

For example, let $X := G \times G$ with the measure $\mu = m_X$, let Σ_X be the Borel σ -algebra of X , and define

$$S^g(x, y) := (gx, y) \quad \text{and} \quad T^g(x, y) := (x, gy).$$

This turns (X, Σ_X, μ, S, T) into a probability $(G \times G)$ -space. It will appear repeatedly below.

Now consider measurable functions $f_1, f_2, f_3 : X \rightarrow [-1, 1]$, and suppose G is highly quasirandom. For Theorems B, C and D we need to estimate the values taken by

$$\text{either } \int f_1 \cdot f_2 S^g \cdot f_3 T^g d\mu \quad \text{or} \quad \int f_1 \cdot f_2 S^g \cdot f_3 S^g T^g d\mu$$

for ‘typical’ group elements g . We ultimately need to do this in case $f_1 = f_2 = f_3 = 1_E$, but allowing more flexibility will be important for the proofs. The strategy for these estimates takes a form that has become well-known in additive combinatorics: each of the functions f_i will be decomposed as

$$f_i = f_i^\circ + f_i^\perp$$

in such a way that the functions f_i^\perp contribute very little to the integrals of interest for ‘most’ group elements g , while the functions f_i° have some extra ‘structure’ which makes the estimate of that integral easier. In such a decomposition, the negligible terms f_i^\perp are often called ‘quasirandom’. Here, ‘most’ group elements will mean those lying in a very large subset of G in the case of Theorem B or C, or those lying in a suitable syndetic subset in the case of Theorem D. A nice discussion of this methodology can be found in [14].

Different decompositions are required for studying naïve and BMZ corners. For the former, the required decompositions are fairly simple, and will be introduced within the proof of Theorem B. For BMZ corners we need a rather more complicated construction, based on the regularity lemma of Frieze and Kannan [12]; this will be explained separately in Subsection 5.1.

3.1 Estimates from probability

The following simple lemma plays the rôle of the classical van der Corput estimate in the present paper. It will be the basis for several other estimates later.

Lemma 1. *Let (Y, Σ, ν) be a probability space, let V be a real or complex Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \|$, and let $y \mapsto u_y$ be a strongly measurable function $Y \rightarrow V$. Let v be a unit vector in V . Then*

$$\int |\langle v, u_y \rangle| \nu(dy) \leq \sqrt{\int \int |\langle u_y, u_{y'} \rangle| \nu(dy) \nu(dy')}.$$

Proof. Define

$$\varphi(y) := \overline{\langle v, u_y \rangle} / |\langle v, u_y \rangle|,$$

using the convention that $\varphi(y) = 1$ if $\langle v, u_y \rangle = 0$. So φ takes values in the unit circle, and more specifically in $\{-1, 1\}$ if V is real. This is a measurable function of y , and

$$\begin{aligned}
\int |\langle v, u_y \rangle| \nu(dy) &= \int \varphi(y) \langle v, u_y \rangle \nu(dy) = \left\langle v, \int \varphi(y) u_y \nu(dy) \right\rangle \\
&\leq \left\| \int \varphi(y) u_y \nu(dy) \right\| = \sqrt{\iint \langle \varphi(y) u_y, \varphi(y') u_{y'} \rangle \nu(dy) \nu(dy')} \\
&\leq \sqrt{\iint |\langle u_y, u_{y'} \rangle| \nu(dy) \nu(dy')}.
\end{aligned}$$

□

We will also need the following very general probabilistic estimate. It can be found as [9, Lemma 1.6], and then [6] cites it for a similar purpose to ours.

Lemma 2. *Let (Y, Σ, ν) be a probability space, let $\Sigma_1, \dots, \Sigma_k$ be σ -subalgebras of Σ , and let f be a bounded non-negative measurable function on Y . Then*

$$\int f \cdot \prod_{i=1}^k \mathbb{E}(f \mid \Sigma_i) d\nu \geq \left(\int f d\nu \right)^{k+1}.$$

□

3.2 Estimates from representation theory

All of the representation theory in this paper concerns either unitary representations on complex Hilbert spaces or orthogonal representations on real Hilbert spaces.

The results we need from representation theory are all standard, and can be found in many textbooks. A good reference for our purposes is [8, Chapters II and III]; that book actually focuses on compact Lie groups, but all the facts we cite from it clearly hold for arbitrary compact groups.

If $\pi : G \curvearrowright V$ is a unitary or orthogonal representation, then P^π denotes the orthogonal projection from V onto the subspace of $\pi(G)$ -fixed vectors.

A unitary (resp. orthogonal) representation is **D -quasirandom** if it has no irreducible subrepresentations of complex (resp. real) dimension less than D . Following [13], the group G itself is D -quasirandom if and only if all its non-identity irreducible unitary representations are D -quasirandom. Quasirandomness will be exploited in this paper by way of the following two lemmas.

Lemma 3. *If G is a D -quasirandom compact group, $\pi : G \curvearrowright V$ is a unitary representation, and $u, v \in V$, then*

$$\|P^{\pi \otimes \pi}(u \otimes v) - P^\pi u \otimes P^\pi v\|_{V \otimes V} \leq D^{-1/2} \|u\|_V \|v\|_V,$$

where $\|\cdot\|_V$ denotes the norm on V and $\|\cdot\|_{V \otimes V}$ denotes the Hilbertian tensor product of that norm on $V \otimes V$.

Proof. This is a routine consequence of Schur’s Lemma: it can be found as [3, Lemma 2]. \square

Unlike the above, the next lemma will be needed for both unitary and orthogonal representations.

Lemma 4. *If G is a D -quasirandom compact group, $\pi : G \curvearrowright V$ is a unitary or orthogonal representation, and $u, v \in V$, then*

$$\int |\langle u, \pi^g v \rangle_V - \langle P^\pi u, P^\pi v \rangle_V|^2 dg \leq D^{-1} \|u\|_V^2 \|v\|_V^2.$$

Proof. In the unitary case, this is [3, Corollary 3] (although the proof given there contained an error, corrected in [2]).

Now suppose that π is an orthogonal representation. Then its complexification $\pi^\mathbb{C} : G \curvearrowright V \otimes_\mathbb{R} \mathbb{C}$ is a unitary representation, which may also be regarded as an orthogonal representation isomorphic to $\pi \oplus \pi$ (see, for instance, [8, Section II.6]). A simple calculation shows that $P^{\pi^\mathbb{C}}$ is simply the complexification of P^π . Hence the desired inequality for π follows from its counterpart for $\pi^\mathbb{C}$. \square

4 Naïve corners

This section proves Theorem B. The key to this proof is the following proposition, which roughly asserts that ‘correlations’ across naïve corners almost vanish unless one starts with functions that have some nontrivial structure.

Let (X, Σ_X, μ, S, T) be as in Section 3.

Proposition 1. *If $f_1, f_2, f_3 : X \rightarrow [-1, 1]$, and either $\mathbb{E}(f_2 | \Sigma_X^S) = 0$ or $\mathbb{E}(f_3 | \Sigma_X^T) = 0$, then*

$$\int \left| \int f_1 \cdot f_2 S^g \cdot f_3 T^g d\mu \right| dg \leq \sqrt{3} D^{-1/4}.$$

Proposition 1 will in turn be deduced from a kind of ‘mixing’ estimate, formulated in the next lemma. Before stating the lemma, it will be helpful to have some notation for representations of product groups. In this section the representations will all be unitary, even though the functions in Proposition 1 are real-valued, since we will need to make another direct appeal to the Schur Orthogonality Relations. Given any two unitary representation $\pi : G \curvearrowright V$ and $\theta : G \curvearrowright W$, one obtains a representation $G \times G \curvearrowright V \otimes W$ by defining

$$(\pi \boxtimes \theta)^{(g,h)} := \pi^g \otimes \theta^h.$$

It is a standard result that $\pi \boxtimes \theta$ is an irreducible $(G \times G)$ -representation if each of π and θ is irreducible, and that all irreducibles of $G \times G$ arise this way [8, Proposition

II.4.14]. (On the other hand, as far as I know there is no standard notation for $\pi \boxtimes \theta$.) It follows at once that if G is D -quasirandom, then so is $G \times G$.

Lemma 5. *Let G be D -quasirandom, let $\rho : G \times G \curvearrowright U$ be a unitary representation, and let $u \in U$. Then*

$$\left\| \int \rho^{(g, g^{-1})} u \, dg - P^\rho u \right\| \leq D^{-1} \|u\|.$$

Proof. By decomposing ρ into irreducibles, it suffices to prove this when ρ is itself a nontrivial irreducible. According to the discussion above, this means that $\rho = \pi \boxtimes \theta$ for some irreducible unitary representations $\pi : G \curvearrowright V$ and $\theta : G \curvearrowright W$, not both the identity.

In this case we will prove that

$$\int \rho^{(g, g^{-1})} (v \otimes w) \, dg = \int (\pi^g v \otimes \theta^{g^{-1}} w) \, dg = \begin{cases} \frac{1}{\dim(V)} (w \otimes v) & \text{if } \pi = \theta \\ 0 & \text{else} \end{cases} \quad (9)$$

for any $v \in V$ and $w \in W$. Since π and θ cannot both be the identity, if they are equal then $\dim(V) \geq D$, by the assumption of quasirandomness. Therefore (9) implies that

$$\left\| \int \rho^{(g, g^{-1})} u \, dg \right\| \leq D^{-1} \|u\| \quad \forall u \in U.$$

To prove (9), it suffices to check the inner products of the two sides of this equation against another element of U , which can also be of tensor product form. Thus the desired equality becomes

$$\begin{aligned} \int \langle \pi^g v \otimes \theta^{g^{-1}} w, v' \otimes w' \rangle \, dg &= \int \langle \pi^g v, v' \rangle \langle \theta^{g^{-1}} w, w' \rangle \, dg \\ &= \begin{cases} \frac{1}{\dim(V)} \langle w, v' \rangle \langle v, w' \rangle & \text{if } \pi = \theta, \\ 0 & \text{else.} \end{cases} \end{aligned}$$

This is now one of the standard Schur Orthogonality Relations: see [8, Theorem II.4.5(ii)]. \square

Corollary 1. *If G is D -quasirandom and $F_1, F_2 \in L^2_{\mathbb{C}}(G \times G)$, then*

$$\left| \int F_1 \cdot \overline{F_2} T^g S^{g^{-1}} \, d\mu - \int F_1 \, d\mu \int \overline{F_2} \, d\mu \right| \leq 2D^{-1/2} \|F_1\|_2 \|F_2\|_2.$$

Remark 1. The use of $\overline{F_2}$ rather than F_2 on the left-hand side here is only for the sake of convenience. With this choice, the integral $\int F_1 \cdot \overline{F_2} T^g S^{g^{-1}} \, d\mu$ is the Hermitian inner product in $L^2_{\mathbb{C}}(G \times G)$, which leads more easily to an application of Lemma 5. \triangleleft

Proof. By replacing each F_i with $F_i - \int F_i \, d\mu$, we may assume that they are both orthogonal to the constant functions. In this case we will show that

$$\int \left| \int F_1 \cdot \overline{F_2 T^g S^{g^{-1}}} d\mu \right|^2 dg \leq 2D^{-1} \|F_1\|^2 \|F_2\|^2,$$

from which the result follows by the Cauchy–Bunyakovski–Schwartz inequality.

This follows using a standard tensor-product trick. Let $\tilde{X} := X \times X$, $\tilde{\mu} := \mu \otimes \mu$, $\tilde{S} := S \times S$ and $\tilde{T} := T \times T$. Then the integral above is equal to

$$\int_G \int_{\tilde{X}} \tilde{F}_1 \cdot \overline{\tilde{F}_2 \tilde{T}^g \tilde{S}^{g^{-1}}} d\tilde{\mu} dg = \int_{\tilde{X}} \tilde{F}_1 \left(\int \overline{\tilde{F}_2 \tilde{T}^g \tilde{S}^{g^{-1}}} dg \right) d\tilde{\mu},$$

where $\tilde{F}_i(x, y, x', y') := F_i(x, y) \overline{F_i(x', y')}$. Applying Lemma 5 to the inner integral on the right here, this is at most

$$\begin{aligned} & \int E(\tilde{F}_1 | \Sigma_{\tilde{X}}^{\tilde{S}, \tilde{T}}) \overline{E(\tilde{F}_2 | \Sigma_{\tilde{X}}^{\tilde{S}, \tilde{T}})} d\tilde{\mu} + D^{-1} \|\tilde{F}_1\|_{L_c^2(\tilde{\mu})} \|\tilde{F}_2\|_{L_c^2(\tilde{\mu})} \\ & \leq \|E(\tilde{F}_1 | \Sigma_{\tilde{X}}^{\tilde{S}, \tilde{T}})\|_{L_c^2(\tilde{\mu})} \|E(\tilde{F}_2 | \Sigma_{\tilde{X}}^{\tilde{S}, \tilde{T}})\|_{L_c^2(\tilde{\mu})} + D^{-1} \|F_1\|^2 \|F_2\|^2. \end{aligned}$$

Finally, in view of the product form of \tilde{F}_i and the fact that $G \times G$ is still D -quasirandom, an appeal to Lemma 3 gives

$$\|E(\tilde{F}_i | \Sigma_{\tilde{X}}^{\tilde{S}, \tilde{T}})\|_{L_c^2(\tilde{\mu})} \leq D^{-1/2} \|F_i\|^2 \quad \text{for } i = 1, 2.$$

Substituting this into the bound above completes the proof.

Proof (Proposition 1). We give the proof in case $E(f_2 | \Sigma_X^S) = 0$, the other case being analogous.

For each $g \in G$, let

$$u_g := f_2 S^g \cdot f_3 T^g.$$

By Lemma 1, it suffices to prove that

$$\iint |\langle u_g, u_{g'} \rangle| dg dg' = \iint |\langle u_g, u_{hg} \rangle| dg dh \leq 3D^{-1/2}.$$

The second integrand here may be re-written as

$$\begin{aligned} |\langle u_g, u_{hg} \rangle| &= \left| \int f_2 S^g \cdot f_2 S^{hg} \cdot f_3 T^g \cdot f_3 T^{hg} d\mu \right| \\ &= \left| \int f_2 \cdot f_2 S^h \cdot f_3 T^g S^{g^{-1}} \cdot f_3 T^{hg} S^{g^{-1}} d\mu \right| \\ &= \left| \int F_{2,h} \cdot F_{3,h} T^g S^{g^{-1}} d\mu \right|, \end{aligned}$$

where $F_{2,h} := f_2 \cdot f_2 S^h$ and $F_{3,h} := f_3 \cdot f_3 T^h$.

Both $F_{2,h}$ and $F_{3,h}$ are real-valued and bounded by 1 in absolute value, so Corollary 1 applies to give

$$\int |\langle u_g, u_{hg} \rangle| dg \leq \left| \int f_2 \cdot f_2 S^h d\mu \right| \left| \int f_3 \cdot f_3 T^h d\mu \right| + 2D^{-1/2}$$

for any h . Now integrating over h , and using again that $|f_3 \cdot f_3 T^h| \leq 1$, the right-hand side above turns into the bound

$$\int \left| \int f_2 \cdot f_2 S^h d\mu \right| dh + 2D^{-1/2} \leq \left(\int \left(\int f_2 \cdot f_2 S^h d\mu \right)^2 dh \right)^{-1/2} + 2D^{-1/2}.$$

By Lemma 4 and the fact that $E(f_2 | \Sigma_X^S) = 0$, this is at most $3D^{-1/2}$. This completes the proof. \square

Proof (Theorem B). We will prove that, for any function $f : G \times G \rightarrow [0, 1]$, the set

$$A := \left\{ g \in G : \iint f(x, y) f(gx, y) f(x, gy) dx dy \geq \left(\int f d\mu \right)^3 - \varepsilon \right\}$$

has measure at least $1 - 2\sqrt{3}D^{-1/4}/\varepsilon$. Applying this to $f = 1_E$ gives Theorem B.

Define new $[-1, 1]$ -valued functions f_2^\perp and f_3^\perp by the decompositions

$$f = E(f | \Sigma_X^S) + f_2^\perp = E(f | \Sigma_X^T) + f_3^\perp.$$

In the present setting, these are the decompositions of $f_2 = f$ and $f_3 = f$ into ‘structured’ and ‘quasirandom’ parts, as promised at the beginning of Section 3. It turns out that no related decomposition $f_1^\circ + f_1^\perp$ is needed here.

Substituting the first decomposition into the middle position of the relevant integral, and then the second decomposition into the last position, we obtain

$$\begin{aligned} & \iint f(x, y) f(gx, y) f(x, gy) dx dy \\ &= \iint f(x, y) E(f | \Sigma_X^S)(gx, y) E(f | \Sigma_X^T)(x, gy) dx dy \\ & \quad + \iint f(x, y) f_2^\perp(gx, y) f(x, gy) dx dy \\ & \quad + \iint f(x, y) E(f | \Sigma_X^S)(gx, y) f_3^\perp(x, gy) dx dy. \end{aligned} \tag{10}$$

Since $E(f | \Sigma_X^S)$ is S -invariant and $E(f | \Sigma_X^T)$ is T -invariant, the first integral on the right-hand side of (10) is equal to

$$\iint f(x, y) E(f | \Sigma_X^S)(x, y) E(f | \Sigma_X^T)(x, y) dx dy$$

for any g , and this is bounded below by $(\int f d\mu)^3$ by Lemma 2. Re-arranging (10), it follows that

$$A \supseteq A' := \left\{ g \in G : \left| \iint f(x, y) f_2^\perp(gx, y) f(x, gy) \, dx \, dy \right| + \left| \iint f(x, y) \mathbb{E}(f \mid \Sigma_X^S)(gx, y) f_3^\perp(x, gy) \, dx \, dy \right| \leq \varepsilon \right\}.$$

On the other hand, since $\mathbb{E}(f_2^\perp \mid \Sigma_X^S) = \mathbb{E}(f_3^\perp \mid \Sigma_X^T) = 0$, two appeals to Proposition 1 give

$$\begin{aligned} & \int \left| \iint f(x, y) f_2^\perp(gx, y) f(x, gy) \, dx \, dy \right| dg \\ & + \int \left| \iint f(x, y) \mathbb{E}(f \mid \Sigma_X^S)(gx, y) f_3^\perp(x, gy) \, dx \, dy \right| dg \leq 2\sqrt{3}D^{-1/4}. \end{aligned}$$

Hence $m_G(A') \geq 1 - 2\sqrt{3}D^{-1/4}/\varepsilon$, by Chebyshev's Inequality, so the proof is complete. \square

5 BMZ corners

This section proves Theorems C and D. The two proofs have much in common. We will explain the overarching structure of both proofs first, and then separate their finer details.

In this section it now makes more sense to work with orthogonal (real) representations than with unitary ones, since we will not need any representation theoretic results beyond Lemma 4.

Again let (X, Σ_X, μ, S, T) be as in Section 3.

5.1 Decomposition into structured and quasirandom functions

We now describe the decompositions into ‘structured’ and ‘quasirandom’ functions appropriate to the analysis of the family of triple forms

$$\int f_1 \cdot f_2 S^g \cdot f_3 S^g T^g \, d\mu, \quad g \in G. \quad (11)$$

The functions f_1 , f_2 and f_3 play different rôles here, so each will need its own notion of ‘quasirandom’ and ‘structured’ summands, even though our ultimate interest is in the case $f_1 = f_2 = f_3 = 1_E$.

As is common in this area, the appropriate notions of ‘quasirandomness’ are measured as smallness in certain norms. We introduce these next.

For a bounded function $f : X \rightarrow \mathbb{R}$, define

$$\|f\|_{\otimes_{1,2}} := \sup \left\{ \int f(x, y)g(x)h(y) \, d\mu : \|g\|_\infty, \|h\|_\infty \leq 1 \right\}.$$

This is a classical construction in Banach space theory: the injective tensor norm $\|\cdot\|_{L^1(m_G) \otimes L^1(m_G)}$. It originates in Schatten's work on tensor products of Banach spaces, where it is referred to as the 'bound cross-norm'. See [17, Section II.3], or [16, Chapter 3] for a more modern treatment. It has now become popular in extremal combinatorics, where it is called the 'box norm'.

Similarly, define

$$\|f\|_{\otimes_{1,12}} := \sup \left\{ \int f(x, y)g(y)h(x^{-1}y) \, d\mu : \|g\|_\infty, \|h\|_\infty \leq 1 \right\}$$

and

$$\|f\|_{\otimes_{12,2}} := \sup \left\{ \int f(x, y)g(x^{-1}y)h(x) \, d\mu : \|g\|_\infty, \|h\|_\infty \leq 1 \right\}.$$

These may also be viewed as injective tensor norms, once one chooses appropriate 'coordinate axes' on $G \times G$.

For general functions f_1, f_2 and f_3 on $G \times G$, we will make use of Frieze and Kannan's weak version of the Szemerédi Regularity Lemma [12] (see also [14, Proposition 2.11] for a formulation closer to the present paper). The next proposition gives the specific instance of this result that we need. Recall that if \mathcal{P} is a partition of a set S and $s \in S$, then $\mathcal{P}(s)$ denotes the cell of \mathcal{P} which contains s .

Proposition 2 (Weak Regularity Lemma). *Given a measurable function $f : X \rightarrow [-1, 1]$, and also $\eta > 0$, there are partitions $\mathcal{P}_{1,1}$ and $\mathcal{P}_{1,12}$ of G , each into at most $\exp((1/\eta)^{O(1)})$ cells, for which the following holds.*

Define a new partition of $G \times G$ cell-wise by setting

$$\mathcal{Q}_1(x, y) := \mathcal{P}_{1,1}(y) \cap \mathcal{P}_{1,12}(x^{-1}y), \quad (12)$$

and let $E_1 : L^\infty(\mu) \rightarrow L^\infty(\mu)$ be the operator of conditional expectation onto \mathcal{Q}_1 . Then

$$\|f - E_1 f\|_{\otimes_{1,12}} \leq \eta.$$

□

This is not the formulation of the Frieze-Kannan Regularity Lemma given in [12] or [14], but it is an easy consequence. The methods of [12] or [14] give instead a function

$$f^\circ(x, y) = \sum_{m=1}^{m_0} \lambda_m h'_m(y) h''_m(x^{-1}y)$$

which approximates f in the sense that

$$\|f - f^\circ\|_{\otimes_{1,12}} \leq \eta,$$

and where

- $m_0 = (1/\eta)^{O(1)}$,
- each $\lambda_m \in [-1, 1]$,
- and each h'_m and h''_m is an indicator function on G .

For our purposes it is important to approximate f by a function that still takes values in $[0, 1]$, hence our preference for approximating by a conditional expectation of f itself. To obtain suitable partitions $\mathcal{P}_{1,1}$ and $\mathcal{P}_{1,12}$ from the function f° above, one simply lets $\mathcal{P}_{1,1}$ be generated by the level sets of all the functions h'_m , $m = 1, 2, \dots, m_0$, and similarly for $\mathcal{P}_{1,12}$ using h''_m .

In recent years, techniques for decomposing a function into structured and quasirandom parts have become quite sophisticated: see [14], for example. A more careful argument than ours might enable one to be more efficient in the above proposition, and perhaps ultimately improve the bound in Theorem C to $D = (1/\varepsilon)^{O(1)}$. However, it is not the purpose of the present paper to explore this kind of enhancement, and we content ourselves with the partitions obtained above.

We will also need two variants of Proposition 2 that are obtained from it by simple changes of variables in $G \times G$. Given f and η as above, there are also partitions $\mathcal{P}_{2,1}$, $\mathcal{P}_{2,2}$, $\mathcal{P}_{3,2}$ and $\mathcal{P}_{3,12}$, all into at most $\exp((1/\eta)^{O(1)})$ cells, for which the following hold: defining two new partitions of $G \times G$ by

$$\mathcal{Q}_2(x, y) := \mathcal{P}_{2,2}(x) \cap \mathcal{P}_{2,1}(y) \quad (13)$$

$$\text{and } \mathcal{Q}_3(x, y) := \mathcal{P}_{3,12}(x^{-1}y) \cap \mathcal{P}_{3,2}(x), \quad (14)$$

and letting $E_i : L^\infty(\mu) \rightarrow L^\infty(\mu)$ be the operator of conditional expectation onto \mathcal{Q}_i for $i = 2, 3$, one has

$$\|f - E_2 f\|_{\otimes_{1,2}} \leq \eta \quad \text{and} \quad \|f - E_3 f\|_{\otimes_{12,2}} \leq \eta.$$

At some points below, we will need to write out the functions $E_i f$, $i = 1, 2, 3$ in terms of more elementary summands. Using the individual cells of \mathcal{Q}_i , one may always express

$$(E_1 f)(x, y) = \sum_{m_1=1}^{M_1} h_{1,1,m_1}(y) h_{1,12,m_1}(x^{-1}y), \quad (15)$$

$$(E_2 f)(x, y) = \sum_{m_2=1}^{M_2} h_{2,1,m_2}(y) h_{2,2,m_2}(x) \quad (16)$$

$$\text{and } (E_3 f)(x, y) = \sum_{m_3=1}^{M_3} h_{3,2,m_3}(x) h_{3,12,m_3}(x^{-1}y), \quad (17)$$

where $M_i = |\mathcal{Q}_i|$ for $i = 1, 2, 3$, and each $h_{\bullet,\bullet,\bullet}$ is a measurable function $G \rightarrow [-1, 1]$.

5.2 Estimates for structured functions

This subsection analyzes the triple form (11) when each f_i is replaced by a structured approximant $E_i f_i$ as given by Proposition 2. For these approximants, we can exert very precise control over that triple form: it turns out that it hardly depends on g at all. This fact will result from the following.

Lemma 6. *Let $h_1, h_2, h_{12} : G \rightarrow [-1, 1]$ be measurable, and suppose that G is D -quasirandom. Then*

$$\left| \iint h_1(x) h_2(y) h_{12}(x^{-1}y) \, dx \, dy - \left(\int h_1 \right) \left(\int h_2 \right) \left(\int h_{12} \right) \right| \leq D^{-1/2}.$$

Proof. This is essentially the implication (v) \implies (iv) in [13, Theorem 4.5]. Let R be the left-action of G on itself. The left-hand side above may be re-written as

$$\begin{aligned} & \left| \int h_1(x) \left(\int h_2(y) h_{12}(x^{-1}y) \, dy - \left(\int h_2 \right) \left(\int h_{12} \right) \right) dx \right| \\ &= \left| \int h_1(x) \left(\langle h_2, h_{12} R^{x^{-1}} \rangle_{L^2(G)} - \left(\int h_2 \right) \left(\int h_{12} \right) \right) dx \right|. \end{aligned}$$

By the Cauchy–Bunyakowski–Schwartz inequality, this is at most

$$\left(\int \left(\langle h_2, h_{12} R^{x^{-1}} \rangle_{L^2(G)} - \left(\int h_2 \right) \left(\int h_{12} \right) \right)^2 dx \right)^{1/2},$$

and Lemma 4 bounds this by $D^{-1/2}$. □

Corollary 2. *Suppose that G is D -quasirandom, and that*

$$\begin{aligned} f_1(x, y) &= h_{1,1}(y) h_{1,12}(x^{-1}y), & f_2(x, y) &= h_{2,1}(y) h_{2,2}(x) \\ & & \text{and } f_3(x, y) &= h_{3,2}(x) h_{3,12}(x^{-1}y) \end{aligned}$$

for some measurable functions $h_{\bullet,\bullet} : G \rightarrow [-1, 1]$. Then the quantity

$$\varphi(g) := \int f_1 \cdot f_2 S^g \cdot f_3 S^g T^g \, d\mu$$

satisfies

$$|\varphi(g) - \varphi(g')| \leq 2D^{-1/2} \quad \forall g, g' \in G.$$

Proof. For this choice of functions f_i , one has

$$\varphi(g) = \iint (h_{1,1} h_{2,1})(y) \cdot (h_{1,12} h_{3,12})(x^{-1}y) \cdot (h_{2,2} h_{3,2})(gx) \, dx \, dy.$$

Applying Lemma 6 to this integral, we find that it lies within $D^{-1/2}$ of

$$\left(\int h_{1,1}h_{2,1}\right)\left(\int h_{1,12}h_{3,12}\right)\left(\int h_{2,2}h_{3,2}\right)$$

for all g . \square

Proposition 3. *Let $\varepsilon > 0$, and let \mathcal{Q}_i , $i = 1, 2, 3$, be partitions as in (12), (13) and (14). Assume that G is D -quasirandom for some*

$$D \geq \frac{16|\mathcal{Q}_1|^2|\mathcal{Q}_2|^2|\mathcal{Q}_3|^2}{\varepsilon^2}. \quad (18)$$

Let the sets E and B be as in Theorem C. Then B contains the set

$$\begin{aligned} C &:= \left\{ g \in G : \mu(E \cap (g^{-1}, e) \cdot E \cap (g^{-1}, g^{-1}) \cdot E) \right. \\ &\quad \left. \geq \int (\mathbf{E}_1 1_E) \cdot (\mathbf{E}_2 1_E) S^g \cdot (\mathbf{E}_3 1_E) S^g T^g d\mu - \varepsilon/2 \right\}. \end{aligned}$$

Proof. According to the decompositions (15)–(17), the quantity

$$\psi(g) := \int (\mathbf{E}_1 1_E) \cdot (\mathbf{E}_2 1_E) S^g \cdot (\mathbf{E}_3 1_E) S^g T^g d\mu$$

is a sum of at most $|\mathcal{Q}_1||\mathcal{Q}_2||\mathcal{Q}_3|$ quantities having the form φ treated by Corollary 2. Therefore, by that corollary, $\psi(g)$ varies by at most

$$2|\mathcal{Q}_1||\mathcal{Q}_2||\mathcal{Q}_3|D^{-1/2} \leq \varepsilon/2$$

as g varies in G . Therefore

$$\mu(E \cap (g^{-1}, e) \cdot E \cap (g^{-1}, g^{-1}) \cdot E) \geq \psi(e) - \varepsilon = \int \mathbf{E}_1 1_E \cdot \mathbf{E}_2 1_E \cdot \mathbf{E}_3 1_E d\mu - \varepsilon$$

for all $g \in C$.

Finally, an appeal to Lemma 2 gives

$$\int \mathbf{E}_1 1_E \cdot \mathbf{E}_2 1_E \cdot \mathbf{E}_3 1_E d\mu \geq \int 1_E \cdot \mathbf{E}_1 1_E \cdot \mathbf{E}_2 1_E \cdot \mathbf{E}_3 1_E d\mu \geq \mu(E)^4.$$

Hence $g \in B$ for all $g \in C$. \square

In the remainder of the proofs of Theorems C and D, we show that the set C from the above corollary is large in the required senses, rather than handle the set B from Theorem C itself.

5.3 Completed proof of Theorem C

In order to use the approximants $E_i f_i$ to the functions f_i in estimating the triple forms (11), we need to convert a bound on the norms such as $\|\cdot\|_{\otimes_{1,2}}$ into some more direct control on those triple forms. That control is given by the next proposition.

Proposition 4. *Let $f_1, f_2, f_3 : X \rightarrow [-1, 1]$ be measurable. Then*

$$\begin{aligned} \int \left| \int f_1 \cdot f_2 S^g \cdot f_3 S^g T^g d\mu \right| dg \\ \leq \sqrt{2D^{-1/4} + \sqrt{\min\{\|f_1\|_{\otimes_{1,12}}, \|f_2\|_{\otimes_{1,2}}, \|f_3\|_{\otimes_{12,2}}\}}}. \end{aligned}$$

This may be viewed as the analog of Proposition 1 for BMZ corners. It will be deduced from the following intermediate estimates.

Lemma 7. *For any $f : X \rightarrow [-1, 1]$, the following inequalities hold:*

$$\int \|E(f \cdot f S^h \mid \Sigma_X^T)\|_2^2 dh \leq D^{-1/2} + \|f\|_{\otimes_{1,2}}, \quad (19)$$

$$\int \|E(f \cdot f (ST)^h \mid \Sigma_X^T)\|_2^2 dh \leq D^{-1/2} + \|f\|_{\otimes_{12,2}}, \quad (20)$$

$$\text{and} \quad \int \|E(f \cdot f (ST)^h \mid \Sigma_X^S)\|_2^2 dh \leq D^{-1/2} + \|f\|_{\otimes_{1,12}}. \quad (21)$$

Proof. We first prove (19). Define a new probability $(G \times G)$ -space as follows: let $Y := G \times G \times G$, let $\nu := m_Y$, and let the two generating actions be

$$\tilde{S}^g(x, y, z) = (gx, y, z) \quad \text{and} \quad \tilde{T}^g(x, y, z) = (x, gy, gz).$$

(In ergodic-theory terms, this is the relative product of two copies of (X, μ, S, T) over Σ_X^T .) Then a simple calculation shows that

$$\|E(f \cdot f S^h \mid \Sigma_X^T)\|_2^2 = \int F \cdot F \tilde{S}^h d\nu = \langle F, F \tilde{S}^h \rangle_{L^2(\nu)},$$

where

$$F(x, y, z) := f(x, y)f(x, z).$$

Let $H := E(F \mid \Sigma_Y^{\tilde{S}})$. Applying Lemma 4 to F and \tilde{S} , it follows that

$$\begin{aligned} \int \|E(f \cdot f S^h \mid \Sigma_X^T)\|_2^2 dh &= \int \langle F, F \tilde{S}^h \rangle_{L^2(\nu)} dh \\ &\leq \int \left| \langle F, F \tilde{S}^h \rangle_{L^2(\nu)} - \langle F, H \rangle_{L^2(\nu)} \right| dh + |\langle F, H \rangle_{L^2(\nu)}| \\ &\leq D^{-1/2} + |\langle F, H \rangle_{L^2(\nu)}|. \end{aligned}$$

On the other hand, H is a function of only the coordinates y and z for a point $(x, y, z) \in Y$, and therefore

$$\begin{aligned} |\langle F, H \rangle_{L^2(\nu)}| &= \left| \iiint f(x, y) f(x, z) H(y, z) \, dx \, dy \, dz \right| \\ &\leq \int \left| \iint f(x, y) \cdot (f(x, z) H(y, z)) \, dx \, dy \right| dz \\ &\leq \int \|f\|_{\otimes_{1,2}} \, dz = \|f\|_{\otimes_{1,2}}. \end{aligned}$$

The proof of (20) is very similar. One uses the same auxiliary $(G \times G)$ -space $(Y, \nu, \tilde{S}, \tilde{T})$ and function F as before, but now one proceeds from the estimate

$$\begin{aligned} \int \|\mathbb{E}(f \cdot f(ST)^h \mid \Sigma_X^T)\|_2^2 \, dh &= \iint F \cdot F(\tilde{S}\tilde{T})^h \, d\nu \, dh \\ &= \int \langle F, F(\tilde{S}\tilde{T})^h \rangle_{L^2(\nu)} \, dh \leq D^{-1/2} + \left| \int F \cdot \mathbb{E}(F \mid \Sigma_Y^{\tilde{S}\tilde{T}}) \, d\nu \right|. \end{aligned}$$

The function $H' := \mathbb{E}(F \mid \Sigma_Y^{\tilde{S}\tilde{T}})$ can be written as a function of only $x^{-1}y$ and $x^{-1}z$ for a point $(x, y, z) \in Y$. Using this, the change of variables $w := x^{-1}z$ gives

$$\begin{aligned} \left| \int F \cdot H' \, d\nu \right| &= \left| \iiint f(x, y) f(x, z) H'(x^{-1}y, x^{-1}z) \, dx \, dy \, dz \right| \\ &= \left| \iiint f(x, y) f(x, xw) H'(x^{-1}y, w) \, dx \, dy \, dw \right| \\ &\leq \int \left| \iint f(x, y) \cdot (f(x, xw) H'(x^{-1}y, w)) \, dx \, dy \right| dw \\ &\leq \int \|f\|_{\otimes_{12,2}} \, dw = \|f\|_{\otimes_{12,2}}. \end{aligned}$$

Finally, inequality (21) is simply (20) with the rôles of S and T reversed. \square

Proof (Proposition 4). We first prove the bound that uses either $\|f_2\|_{\otimes_{1,2}}$ or $\|f_3\|_{\otimes_{12,2}}$. Let $u_g := f_2 S^g \cdot f_3 S^g T^g$. By Lemma 1, it suffices to prove that

$$\begin{aligned} \iint |\langle u_g, u_{g'} \rangle| \, dg \, dg' &= \iint |\langle u_g, u_{hg} \rangle| \, dg \, dh \\ &\leq 2D^{-1/4} + \sqrt{\min\{\|f_2\|_{\otimes_{1,2}}, \|f_3\|_{\otimes_{12,2}}\}}. \end{aligned}$$

For any $g, h \in G$, one has

$$\begin{aligned}
|\langle u_g, u_{hg} \rangle| &= \left| \int f_2 S^g \cdot f_2 S^{hg} \cdot f_3 S^g T^g \cdot f_3 S^{hg} T^{hg} d\mu \right| \\
&= \left| \int f_2 \cdot f_2 S^h \cdot f_3 T^g \cdot f_3 S^h T^{hg} d\mu \right| \\
&= |\langle F_h, F'_h T^g \rangle|,
\end{aligned}$$

where $F_h := f_2 \cdot f_2 S^h$ and $F'_h := f_3 \cdot f_3 (ST)^h$, both of which take values in $[-1, 1]$. Therefore Lemma 4 and the Cauchy–Bunyakowski–Schwartz Inequality give

$$\begin{aligned}
\iint |\langle u_g, u_{hg} \rangle| dg dh &\leq \int |\langle \mathbb{E}(F_h | \Sigma_X^T), \mathbb{E}(F'_h | \Sigma_X^T) \rangle| dh + D^{-1/2} \\
&\leq \int \|\mathbb{E}(F_h | \Sigma_X^T)\|_2 \|\mathbb{E}(F'_h | \Sigma_X^T)\|_2 dh + D^{-1/2} \\
&\leq \sqrt{\int \|\mathbb{E}(F_h | \Sigma_X^T)\|_2^2 dh \cdot \int \|\mathbb{E}(F'_h | \Sigma_X^T)\|_2^2 dh} + D^{-1/2}.
\end{aligned}$$

Each of the integrals inside this last square root is certainly at most 1. Therefore, by Lemma 7, the last line above may be bounded by

$$\begin{aligned}
\text{either } \sqrt{\|f_2\|_{\otimes_{1,2}}^2 + D^{-1/2}} + D^{-1/2} &\leq \sqrt{\|f_2\|_{\otimes_{1,2}}^2} + 2D^{-1/4} \\
\text{or } \sqrt{\|f_3\|_{\otimes_{12,2}}^2 + D^{-1/2}} + D^{-1/2} &\leq \sqrt{\|f_3\|_{\otimes_{12,2}}^2} + 2D^{-1/4}.
\end{aligned}$$

Finally, the proof of the bound using $\|f_1\|_{\otimes_{1,12}}$ is exactly analogous to the case of $\|f_3\|_{\otimes_{12,2}}$ once one makes the substitution $g' := g^{-1}$ to write

$$\int \left| \int f_1 \cdot f_2 S^g \cdot f_3 S^g T^g d\mu \right| dg = \int \left| \int f_3 \cdot f_2 T^{g'} \cdot f_1 S^{g'} T^{g'} d\mu \right| dg'.$$

□

Proof (Theorem C). Let $f := 1_E$ and $\eta := \varepsilon^8/(4 \cdot 6^4)$. Then there are certainly values of D satisfying (6) for which

$$\sqrt{2D^{-1/4} + \sqrt{\eta}} \leq \varepsilon^2/6. \tag{22}$$

For this function f and error tolerance η , let the partitions \mathcal{Q}_i , $i = 1, 2, 3$, and corresponding operators \mathbb{E}_i be given by Proposition 2 and its two variants. Then $|\mathcal{Q}_i| \leq \exp(2(1/\eta)^{O(1)})$ for each i , and so there are values of D satisfying (6) for which (18) also holds. Therefore Proposition 3 applies, and so it suffices to show that the set C from that proposition has measure at least $1 - \varepsilon$. We will in fact prove that

$$\int \left| \int f \cdot f S^g \cdot f S^g T^g d\mu - \int (\mathbb{E}_1 f) \cdot (\mathbb{E}_2 f) S^g \cdot (\mathbb{E}_3 f) S^g T^g d\mu \right| dg \leq \varepsilon^2/2.$$

The desired lower bound on $m_G(C)$ follows from this by Chebyshev's Inequality.

Let $f_i^\perp = f - E_i f$ for each i . By the triangle inequality, the above estimate is a consequence of the following:

$$\begin{aligned} & \int \left| \int f \cdot f S^g \cdot f_3^\perp S^g T^g d\mu \right| dg \leq \varepsilon^2/6, \\ & \int \left| \int f \cdot f_2^\perp S^g \cdot (E_3 f) S^g T^g d\mu \right| dg \leq \varepsilon^2/6 \\ \text{and} \quad & \int \left| \int f_1^\perp \cdot (E_2 f) S^g \cdot (E_3 f) S^g T^g d\mu \right| dg \leq \varepsilon^2/6. \end{aligned}$$

These are all now implied by Proposition 4, together with (22). \square

5.4 Anti-neighbourhoods and syndeticity

The proof of Theorem D will be based on having a large supply of fairly syndetic subsets of a quasirandom group ready to hand. These subsets will be obtained from a simple construction in terms of representations.

Let $\pi : G \curvearrowright V$ be an orthogonal representation, let $u, v \in V$, and let $\varepsilon > 0$. If $P^\pi = 0$ (that is, π contains no copy of the identity representation), then let

$$A(\pi, u, v, \varepsilon) := \{g \in G : |\langle u, \pi^g v \rangle| < \varepsilon\},$$

and call this the (π, u, v, ε) -**anti-neighbourhood**. For general π , let

$$A(\pi, u, v, \varepsilon) := \{g \in G : |\langle u, \pi^g v \rangle - \langle P^\pi u, P^\pi v \rangle| < \varepsilon\}.$$

If G is D -quasirandom for some large D , and u, v are unit vectors, then the corresponding anti-neighbourhoods are quite large: Lemma 4 and Chebyshev's Inequality imply that

$$m_G(A(\pi, u, v, \varepsilon)) \geq 1 - \frac{D^{-1}}{\varepsilon^2}. \quad (23)$$

This is very intuitive. If π is irreducible with large dimension d , then the orbit points $\pi^g v$ should be fairly evenly spread around the high-dimensional unit sphere in V , and so most of them will be nearly orthogonal to any fixed direction u .

The present section shows that, if G is highly quasirandom, then anti-neighbourhoods are also fairly syndetic. Moreover, one can intersect a controlled number of anti-neighbourhoods, and that intersection is still fairly syndetic. This is not implied solely by the largeness of those intersections, but it will follow from some simple inner-product estimates. In the next section, the syndeticity of the set in Theorem D will be proved by showing that it contains such an intersection of anti-neighbourhoods.

Let us begin with a rough sketch of why a single anti-neighbourhood should be fairly syndetic, before carefully proving the result we need about intersections. The key point is that, if d is very large and we choose a moderately large number of group elements h_1, \dots, h_k independently at random, then with high probability the image-points $\pi^{h_i}u$ will all be nearly orthogonal to one another. The intuition here is the same as that above: each of the random points $\pi^{h_i}u$ should be fairly evenly distributed around the high-dimensional unit sphere of V , independently of the others. However, having obtained some h_i 's for which the vectors $\pi^{h_i}u$ are all nearly orthogonal, this then forces any other unit vector to be nearly orthogonal to at least some of them. In particular, for any $g \in G$, there is an i such that $\pi^{h_i}u$ and $\pi^g v$ are nearly orthogonal. This implies that u is nearly orthogonal to $\pi^{h_i^{-1}g}v$, and hence that g is in $h_i \cdot A(\pi, u, v, \varepsilon)$ for some small ε . Crucially, after fixing the right choice of h_1, \dots, h_k , this argument works for every $g \in G$.

Such ‘near orthogonality’ will be deduced using the following.

Lemma 8. *Let V be a real Hilbert space and let v and u^1, \dots, u^m be unit vectors in V . Suppose that*

$$|\langle u^i, u^j \rangle| \leq 1/m^2 \quad \text{whenever } i \neq j.$$

Then

$$\sum_{i \leq m} \langle v, u^i \rangle^2 \leq 2.$$

Proof. Let $a_i := \langle v, u^i \rangle$ for each i , so one always has $|a_i| \leq 1$, and let

$$w := \sum_{i \leq m} a_i u^i.$$

The assumed inequalities give

$$\|w\|^2 = \sum_{i,j \leq m} a_i a_j \langle u^i, u^j \rangle \leq \sum_{i \leq m} a_i^2 + (m^2 - m)/m^2 \leq \sum_{i \leq m} a_i^2 + 1,$$

and hence

$$\sum_{i \leq m} a_i^2 = \langle v, w \rangle \leq \|v\| \|w\| \leq \sqrt{\sum_{i \leq m} a_i^2 + 1}.$$

This implies that $\sum_{i \leq m} a_i^2 \leq 2$. □

In order to study intersections of anti-neighbourhoods, we will actually need the following crude corollary which concerns several Hilbert spaces simultaneously.

Corollary 3. *Let V_1, \dots, V_k be real Hilbert spaces, and let v_ℓ and $u_\ell^1, \dots, u_\ell^m$ be unit vectors in V_ℓ for each ℓ . Suppose that*

$$|\langle u_\ell^i, u_\ell^j \rangle| \leq 1/m^2 \quad \text{whenever } \ell \leq k, i \neq j.$$

Then there is some $i \leq m$ such that $|\langle v_\ell, u_\ell^i \rangle| \leq \sqrt{2k/m}$ for all $\ell \leq k$.

Proof. Summing the inequalities proved in the preceding lemma gives

$$\sum_{i \leq m} \left(\sum_{\ell \leq k} \langle v_\ell, u_\ell^i \rangle^2 \right) \leq 2k.$$

□

Corollary 4. *Let $\eta > 0$ and $k \geq 1$, and set $K := \lceil 2k/\eta^2 + 1 \rceil$ and $D := K^6 k + 1$. Let $\pi_\ell : G \curvearrowright V_\ell$ for $\ell = 1, 2, \dots, k$ be orthogonal representations that are all D -quasirandom, and let $u_\ell, v_\ell \in V_\ell$ be unit vectors for each ℓ . Then*

$$A := \bigcap_{\ell=1}^k A(\pi_\ell, u_\ell, v_\ell, \eta)$$

is K -syndetic.

Proof. Let

$$A' := \bigcap_{\ell=1}^k A(\pi_\ell, u_\ell, u_\ell, 1/K^2).$$

Let h_1, \dots, h_K be a K -tuple of elements of G drawn at random from the measure $m_G^{\otimes K}$. Then the estimate (23) and a first-moment bound give

$$\begin{aligned} & m_G^{\otimes K} \{ h_i^{-1} h_j \in A' \mid \forall i \neq j \text{ in } \{1, 2, \dots, K\} \} \\ & \geq 1 - \sum_{i \neq j} \sum_{\ell=1}^k m_G(G \setminus A(\pi_\ell, u_\ell, u_\ell, 1/K^2)) \geq 1 - K^6 k D^{-1} > 0. \end{aligned}$$

This implies that there exists a K -tuple h_1, \dots, h_K in G witnessing the above event, hence such that

$$|\langle \pi_\ell^{h_i} u_\ell, \pi_\ell^{h_j} u_\ell \rangle| \leq 1/K^2 \quad \text{whenever } \ell \leq k, i \neq j.$$

Therefore, for any $g \in G$, Corollary 3 promises some $i \leq K$ for which

$$|\langle \pi_\ell^{h_i} u_\ell, \pi_\ell^g v_\ell \rangle| = |\langle u_\ell, \pi_\ell^{h_i^{-1}g} v_\ell \rangle| \leq \sqrt{2k/K} < \eta \quad \forall \ell \leq k,$$

so $g \in \{h_1, \dots, h_K\} \cdot A$. □

5.5 Completed proof of Theorem D

The next step is the following rather technical proposition.

Proposition 5. *Let $\varepsilon > 0$ and $n \in \mathbb{N}$. Set $k := \lceil 4/\varepsilon^2 \rceil$, and now set $\eta := 1/(3k)^8$. Suppose that G is D -quasirandom for some $D > 4k^4/\eta^4$.*

For $\ell = 2, 3$, let f_1^ℓ , f_2^ℓ and f_3^ℓ be three $[-1, 1]$ -valued functions, and suppose that

$$\|f_2^2\|_{\otimes_{1,2}} \leq \eta \quad \text{and} \quad \|f_3^3\|_{\otimes_{12,2}} \leq \eta.$$

Finally, let

$$C_\ell := \left\{ g \in G : \left| \int f_1^\ell \cdot f_2^\ell S^g \cdot f_3^\ell S^g T^g d\mu \right| < \varepsilon \right\} \quad \text{for } \ell = 2, 3.$$

Then there are elements $h_1, \dots, h_k \in G$ and some auxiliary $[-1, 1]$ -valued functions

$$F_{2,i,j}^2, \quad F_{3,i,j}^2, \quad F_{2,i,j}^3 \quad \text{and} \quad F_{3,i,j}^3 \quad \text{for } 1 \leq i < j \leq k$$

such that the set

$$E := \bigcap_{1 \leq i < j \leq k} (A(T, F_{2,i,j}^2, F_{3,i,j}^2, \eta) \cap A(T, F_{2,i,j}^3, F_{3,i,j}^3, \eta))$$

satisfies

$$E \subseteq \{h_1^{-1}, \dots, h_k^{-1}\} \cdot (C_2 \cap C_3).$$

Remark 2. Letting $m := \lceil 2k^2/\eta^2 + 1 \rceil$, the set E above is m -syndetic by Corollary 4 provided D large enough, and so the above conclusion implies that the intersection $C_2 \cap C_3$ is (mk) -syndetic. However, this fact alone is not quite what we need for the proof of Theorem D: ultimately, that will require the syndeticity of the smaller intersection $C_1 \cap C_2 \cap C_3$ for some other set C_1 . In order to prove that, it will be important to have an explicit ‘witness’ to the syndeticity of $C_2 \cap C_3$ in the form of an intersection of anti-neighbourhoods, such as E above. This is why the above proposition is formulated as it is. \triangleleft

Proof. We will prove that if h_1, \dots, h_k are chosen independently at random from the Haar measure m_G , then with positive probability one obtains a tuple for which the remaining objects required by the proposition also exist.

Step 1. For such a random choice of h_1, \dots, h_k , each difference $h_j h_i^{-1}$ for $i \neq j$ is also a uniform random element of G , and so Lemma 7 gives the estimates

$$\int \int \dots \int \mathbb{E} \|f_2^2 \cdot f_2^2 S^{h_j h_i^{-1}} \mid \Sigma_X^T\|_2^2 dh_1 \dots dh_{k-1} dh_k \leq D^{-1/2} + \eta$$

and

$$\int \int \dots \int \mathbb{E} \|f_3^3 \cdot f_3^3 (ST)^{h_j h_i^{-1}} \mid \Sigma_X^T\|_2^2 dh_1 \dots dh_{k-1} dh_k \leq D^{-1/2} + \eta$$

for each $i \neq j$.

Our assumptions imply that $D^{-1/2} < \eta$, and so the right-hand sides above are all less than 2η . Therefore, if h_1, \dots, h_k are chosen randomly as described above, then, by Chebyshev’s Inequality and a first-moment bound, the event that

$$\begin{aligned} \|\mathbb{E}(f_2^2 \cdot f_2^2 S^{h_j h_i^{-1}} \mid \Sigma_X^T)\|_2^2 &< \sqrt{2\eta} \quad \text{for all } i \neq j \\ \text{and } \|\mathbb{E}(f_3^3 \cdot f_3^3 (ST)^{h_j h_i^{-1}} \mid \Sigma_X^T)\|_2^2 &< \sqrt{2\eta} \quad \text{for all } i \neq j \end{aligned} \quad (24)$$

has probability at least

$$1 - 2k^2 \sqrt{2\eta} = 1 - 2\sqrt{2}k^2/(3k)^4 > 0.$$

Therefore there exists a tuple $h_1, \dots, h_k \in G$ for which the inequalities in (24) all hold simultaneously.

Step 2. We now define the required auxiliary functions as follows:

$$\begin{aligned} F_{2,i,j}^2 &:= f_2^2 S^{h_i} \cdot f_2^2 S^{h_j} & \text{and } F_{3,i,j}^2 &:= f_3^2 S^{h_i} T^{h_i} \cdot f_3^2 S^{h_j} T^{h_j} \\ F_{2,i,j}^3 &:= f_2^3 S^{h_i} \cdot f_2^3 S^{h_j} & \text{and } F_{3,i,j}^3 &:= f_3^3 S^{h_i} T^{h_i} \cdot f_3^3 S^{h_j} T^{h_j} \end{aligned}$$

whenever $1 \leq i < j \leq k$.

Step 3. Having chosen those auxiliary functions, let E be as in the statement of the proposition. Suppose that $g \in E$. We must show that $h_i g \in C_2 \cap C_3$ for some $i \leq k$.

For each $g \in G$ and $\ell = 2, 3$, define

$$u_g^\ell := f_2^\ell S^g \cdot f_3^\ell S^g T^g.$$

For any $i < j$ in $\{1, 2, \dots, k\}$ and $\ell = 2, 3$, we have

$$\begin{aligned} |\langle u_{h_i g}^\ell, u_{h_j g}^\ell \rangle| &= \left| \int f_2^\ell S^{h_i g} \cdot f_2^\ell S^{h_j g} \cdot f_3^\ell S^{h_i g} T^{h_i g} \cdot f_3^\ell S^{h_j g} T^{h_j g} \, d\mu \right| \\ &= \left| \int f_2^\ell S^{h_i} \cdot f_2^\ell S^{h_j} \cdot f_3^\ell S^{h_i} T^{h_i g} \cdot f_3^\ell S^{h_j} T^{h_j g} \, d\mu \right| \\ &= \left| \int F_{2,i,j}^\ell \cdot F_{3,i,j}^\ell T^g \, d\mu \right|. \end{aligned}$$

Since $g \in E$, in particular $g \in A(T, F_{2,i,j}^\ell, F_{3,i,j}^\ell, \eta)$, and so the above is at most

$$\left| \int \mathbb{E}(F_{2,i,j}^\ell \mid \Sigma_X^T) \mathbb{E}(F_{3,i,j}^\ell \mid \Sigma_X^T) \, d\mu \right| + \eta.$$

If $\ell = 2$ then this expression is bounded by

$$\begin{aligned} \|\mathbb{E}(F_{2,i,j}^2 \mid \Sigma_X^T)\|_2 + \eta &= \|\mathbb{E}(f_2^2 S^{h_i} \cdot f_2^2 S^{h_j} \mid \Sigma_X^T)\|_2 + \eta \\ &= \|\mathbb{E}(f_2^2 \cdot f_2^2 S^{h_j h_i^{-1}} \mid \Sigma_X^T)\|_2 + \eta < \sqrt[4]{2\eta} + \eta < 3\sqrt[4]{\eta} = 1/k^2, \end{aligned}$$

by (24). Similarly, if $\ell = 3$ then it is bounded by

$$\|\mathbb{E}(F_{3,i,j}^3 \mid \Sigma_X^T)\|_2 + \eta = \|\mathbb{E}(f_3^3 \cdot f_3^3 (ST)^{h_j h_i^{-1}} \mid \Sigma_X^T)\|_2 + \eta < 1/k^2.$$

Thus, we have shown that

$$|\langle u_{h_{ig}}^\ell, u_{h_{jg}}^\ell \rangle| < 1/k^2 \quad \text{whenever } i \neq j, \ell = 2, 3.$$

We may therefore apply Corollary 3 to the inner products

$$\langle f_1^\ell, u_{h_{ig}}^\ell \rangle, \quad i = 1, 2, \dots, k, \quad \ell = 2, 3$$

to conclude that, for any $g \in E$, there is at least one $i \leq k$ for which

$$|\langle f_1^2, u_{h_{ig}}^2 \rangle| \leq \sqrt{4/k} < \varepsilon \quad \text{and} \quad |\langle f_1^3, u_{h_{ig}}^3 \rangle| < \varepsilon.$$

For this choice of i one has $h_{ig} \in C_2 \cap C_3$, as required. \square

Proof (Theorem D). Let $f := 1_E$. As for Theorem C, this proof will be based on three different decompositions of f as given by Proposition 2. However, a new complication here is that the partition \mathcal{Q}_1 will need to be chosen after \mathcal{Q}_2 and \mathcal{Q}_3 , and considerably finer than those others.

In the course of the proof, we will meet three points at which we require a lower bound on D . All of these lower bounds will be satisfied for some D as in (7), so there is a choice of D of the form in (7) for which the whole proof can be carried out.

Step 1. Set $k := \lceil 36/\varepsilon^2 \rceil$ and $\eta := 1/(3k)^8$. Let \mathcal{Q}_2 and \mathcal{Q}_3 be partitions as given by the two variants of Proposition 2 for this error tolerance η , and consider the resulting decompositions

$$f = f_2^\perp + E_2 f = f_3^\perp + E_3 f.$$

Let

$$C_2 := \left\{ g \in G : \left| \int f \cdot f_2^\perp S^g \cdot f S^g T^g d\mu \right| < \varepsilon/3 \right\}$$

and

$$C_3 := \left\{ g \in G : \left| \int f \cdot (E_2 f) S^g \cdot f_3^\perp S^g T^g d\mu \right| < \varepsilon/3 \right\}.$$

Given the above choice of k and η , there is a D as in (7) which satisfies $D > 4k^4/\eta^4$ (indeed, at this point (7) leaves vastly more room than we need). Therefore Proposition 5 applies to give $h_1, \dots, h_k \in G$ and some auxiliary $[-1, 1]$ -valued functions

$$F_{2,i,j}^2, \quad F_{3,i,j}^2, \quad F_{2,i,j}^3 \quad \text{and} \quad F_{3,i,j}^3 \quad \text{for } 1 \leq i < j \leq k$$

such that $E \subseteq \{h_1^{-1}, \dots, h_k^{-1}\} \cdot (C_2 \cap C_3)$, where

$$E = \bigcap_{1 \leq i < j \leq k} (A(T, F_{2,i,j}^2, F_{3,i,j}^2, \eta) \cap A(T, F_{2,i,j}^3, F_{3,i,j}^3, \eta)).$$

Let $M_i := |\mathcal{Q}_i|$ for $i = 2, 3$.

Step 2. Now set $\eta' := \varepsilon/6M_2M_3$, let \mathcal{Q}_1 be as given by Proposition 2 for this error tolerance η' , and consider the decomposition

$$f = f_1^\perp + E_1 f.$$

Let

$$C_1 := \left\{ g \in G : \left| \int f_1^\perp \cdot (E_2 f) S^g \cdot (E_3 f) S^g T^g d\mu \right| < \varepsilon/3 \right\}.$$

Observe that

$$\eta' \geq (\varepsilon/6) \exp(-4(1/\eta)^{O(1)}) \geq \exp(-(1/\varepsilon)^{O(1)})$$

for $\varepsilon \in (0, 1/2)$. Therefore

$$\frac{16|\mathcal{Q}_1|^2|\mathcal{Q}_2|^2|\mathcal{Q}_3|^2}{\varepsilon^2} \leq \frac{16 \exp(4(1/\eta')^{O(1)} + 8(1/\eta)^{O(1)})}{\varepsilon^2} \leq \exp(\exp((1/\varepsilon)^{O(1)}))$$

for all $\varepsilon \in (0, 1/2)$, and so there are values of D satisfying (7) for which (18) holds for these partitions \mathcal{Q}_1 , \mathcal{Q}_2 and \mathcal{Q}_3 . (This is the only point at which we need the double exponential in (7).) Therefore Proposition 3 applies, and so it suffices to show that the set C from that proposition is K -syndetic. Moreover, that set C clearly contains $C_1 \cap C_2 \cap C_3$, so it suffices to show that this triple intersection is K -syndetic.

Step 3. We now make use of the decompositions (16) and (17). Substituting these into the integral that appears inside the definition of C_1 , we obtain

$$\sum_{m_2, m_3} \iint (f_1^\perp(x, y) h_{2,1,m_2}(y) h_{3,12,m_3}(x^{-1}y)) (h_{2,2,m_2} h_{3,2,m_3})(gx) dx dy. \quad (25)$$

Let

$$\psi_{m_2, m_3}(x, y) := f_1^\perp(x, y) h_{2,1,m_2}(y) h_{3,12,m_3}(x^{-1}y),$$

let

$$\varphi_{m_2, m_3}(x, y) := (h_{2,2,m_2} h_{3,2,m_3})(x)$$

(so φ depends only nominally on y), and let

$$E' := \bigcap_{m_2, m_3} A(S, \psi_{m_2, m_3}, \varphi_{m_2, m_3}, \eta').$$

We will now show that $E' \subseteq C_1$, so suppose that $g \in E'$. Then the definition of E' and the expression (25) give

$$\begin{aligned}
\left| \int f_1^\perp \cdot (E_2 f) S^g \cdot (E_3 f) S^g T^g d\mu \right| &\leq \sum_{m_2, m_3} \left| \int \psi_{m_2, m_3} \cdot \varphi_{m_2, m_3} S^g d\mu \right| \\
&\leq \sum_{m_2, m_3} \left(\left| \int \psi_{m_2, m_3} d\mu \right| \left| \int \varphi_{m_2, m_3} d\mu \right| + \eta' \right) \\
&\leq \sum_{m_2, m_3} \left| \int \psi_{m_2, m_3} d\mu \right| + \varepsilon/6.
\end{aligned}$$

Substituting from the definition of ψ_{m_2, m_3} , this is

$$\begin{aligned}
\sum_{m_2, m_3} \left| \iint f_1^\perp(x, y) h_{2,1, m_2}(y) h_{3,12, m_3}(x^{-1}y) dx dy \right| + \varepsilon/6 \\
\leq M_2 M_3 \|f_1^\perp\|_{\otimes_{1,12}} + \varepsilon/6 \leq M_2 M_3 \eta' + \varepsilon/6 = \varepsilon/3,
\end{aligned}$$

so $g \in C_1$.

Step 4. Finally, letting

$$E'' = \bigcap_{i=1}^k h_i^{-1} E' = \bigcap_{i=1}^k \bigcap_{m_2, m_3} A(S, \psi_{m_2, m_3}, \varphi_{m_2, m_3} S^{h_i}, \eta'),$$

it follows that

$$E \cap E'' \subseteq \{h_1^{-1}, \dots, h_k^{-1}\} \cdot (C_1 \cap C_2 \cap C_3) \subseteq \{h_1^{-1}, \dots, h_k^{-1}\} \cdot C.$$

On the other hand,

$$\begin{aligned}
E \cap E'' &= \bigcap_{1 \leq i < j \leq k} (A(T, F_{2,i,j}^2, F_{3,i,j}^2, \eta) \cap A(T, F_{2,i,j}^3, F_{3,i,j}^3, \eta)) \\
&\quad \cap \bigcap_{i=1}^k \bigcap_{m_2, m_3} A(S, \psi_{m_2, m_3}, \varphi_{m_2, m_3} S^{h_i}, \eta').
\end{aligned}$$

Since $\varepsilon < 1/2$, this is an intersection of at most

$$2k^2 + kM_2M_3 \leq 2k^2 + k \exp(4(1/\eta)^{O(1)}) \leq \exp((1/\varepsilon)^{O(1)})$$

anti-neighbourhoods for $\varepsilon \in (0, 1/2)$. On the other hand,

$$\eta, \eta' \geq \exp(-(1/\varepsilon)^{O(1)}).$$

Let

$$K_0 := \lceil 2(2k^2 + kM_2M_3) / \min\{\eta, \eta'\}^2 + 1 \rceil \leq \exp((1/\varepsilon)^{O(1)}).$$

There is a D satisfying (7) for which

$$D \geq K_0^6(2k^2 + kM_2M_3) + 1.$$

Therefore Corollary 4 can be applied to deduce that this intersection is still K_0 -syndetic. Therefore C is K -syndetic for

$$K = K_0k = \exp((1/\varepsilon)^{O(1)}),$$

as required. \square

Remark 3. The above proof uses three different decompositions of f , one for each the three positions in the triple form (11). However, these decompositions do not have equal status: the decomposition $f = f_1^\perp + E_1f$ (corresponding to the first position) requires a much finer partition \mathcal{Q}_1 , and depends on having already obtained the partitions \mathcal{Q}_2 and \mathcal{Q}_3 corresponding to the second and third positions. This contrasts with the proof of Theorem C, where the three positions in the triple form (11) have roughly equal status.

The reason for this difference can be seen in the proof of Proposition 4. First, we proved the bound that uses the norm $\|f_2\|_{\otimes_{1,2}}$ or $\|f_3\|_{\otimes_{12,2}}$; then, the bound that uses $\|f_1\|_{\otimes_{1,12}}$ was obtained by the same argument upon changing variables to $g' := g^{-1}$. However, this change of variables converts left-syndeticity to right-syndeticity, so we cannot use it in the same way to analyze the set C_1 in the proof of Theorem D. Instead, we must first replace f with its structured approximations E_2f (in the second position) and E_3f (in the third position), and then use the special structure of those approximations to analyze the contribution of f in the first position of the triple form, without changing variables.

This discussion also suggests why our methods fail to answer Question 1 from the Introduction (about syndeticity in the setting of Theorem B). The current version of Theorem B involves an estimate of the triple form for naïve corners which uses an integral of the kind appearing in Corollary 1. That integral features both T^g and $S^{g^{-1}}$. Once again, the appearance of g^{-1} converts left-syndetic sets into right-syndetic sets, and so it is not clear how to obtain control on this integral on any particular anti-neighbourhood. \triangleleft

Acknowledgements Research supported by a fellowship from the Clay Mathematics Institute. I am grateful to Vitaly Bergelson for sharing [6] with me, to Ben Green for pointing me to the references [11, 15], to Sean Eberhard for pointing me to the reference [20], and to Julia Wolf for suggesting some useful clarifications.

References

1. Ajtai, M., Szemerédi, E.: Sets of lattice points that form no squares. *Stud. Sci. Math. Hungar.* **9**, 9–11 (1975) (1974)
2. Austin, T.: Quantitative Equidistribution for Certain Quadruples in Quasi-Random Groups: Erratum. To appear, *Combin. Probab. Comput.*

3. Austin, T.: Quantitative Equidistribution for Certain Quadruples in Quasi-Random Groups. *Combin. Probab. Comput.* **24**(2), 376–381 (2015). DOI 10.1017/S0963548314000492. URL <http://dx.doi.org/10.1017/S0963548314000492>
4. Bergelson, V., McCutcheon, R.: Central sets and a non-commutative Roth theorem. *Amer. J. Math.* **129**(5), 1251–1275 (2007). DOI 10.1353/ajm.2007.0031. URL <http://dx.doi.org/10.1353/ajm.2007.0031>
5. Bergelson, V., McCutcheon, R., Zhang, Q.: A Roth theorem for amenable groups. *Amer. J. Math.* **119**(6), 1173–1211 (1997)
6. Bergelson, V., Robertson, D., Zorin-Kranich, P.: Triangles in Cartesian squares of quasirandom groups. Preprint, available online at [arXiv.org: 1410.5385](http://arxiv.org/1410.5385)
7. Bergelson, V., Tao, T.: Multiple recurrence in quasirandom groups. *Geom. Funct. Anal.* **24**(1), 1–48 (2014). DOI 10.1007/s00039-014-0252-0. URL <http://dx.doi.org/10.1007/s00039-014-0252-0>
8. Bröcker, T., tom Dieck, T.: Representations of Compact Lie Groups. Springer (1985)
9. Chu, Q.: Convergence of weighted polynomial multiple ergodic averages. *Proc. Amer. Math. Soc.* **137**, 1363–1369 (2009)
10. Chu, Q., Zorin-Kranich, P.: Lower bound in the Roth Theorem for amenable groups. To appear, *Ergodic Theory Dynam. Systems*
11. Erdős, P., Straus, E.G.: How abelian is a finite group? *Linear and Multilinear Algebra* **3**(4), 307–312 (1975/76)
12. Frieze, A., Kannan, R.: Quick approximation to matrices and applications. *Combinatorica* **19**(2), 175–220 (1999). DOI 10.1007/s004930050052. URL <http://dx.doi.org/10.1007/s004930050052>
13. Gowers, W.T.: Quasirandom groups. *Combin. Probab. Comput.* **17**(3), 363–387 (2008). DOI 10.1017/S0963548307008826. URL <http://dx.doi.org/10.1017/S0963548307008826>
14. Gowers, W.T.: Decompositions, approximate structure, transference, and the Hahn-Banach theorem. *Bull. Lond. Math. Soc.* **42**(4), 573–606 (2010). DOI 10.1112/blms/bdq018. URL <http://dx.doi.org/10.1112/blms/bdq018>
15. Pyber, L.: How abelian is a finite group? In: The mathematics of Paul Erdős, I, *Algorithms Combin.*, vol. 13, pp. 372–384. Springer, Berlin (1997). DOI 10.1007/978-3-642-60408-9_27. URL http://dx.doi.org/10.1007/978-3-642-60408-9_27
16. Ryan, R.A.: Introduction to tensor products of Banach spaces. Springer Monographs in Mathematics. Springer-Verlag London, Ltd., London (2002). DOI 10.1007/978-1-4471-3903-4. URL <http://dx.doi.org/10.1007/978-1-4471-3903-4>
17. Schatten, R.: A Theory of Cross-Spaces. *Annals of Mathematics Studies*, no. 26. Princeton University Press, Princeton, N. J. (1950)
18. Shkredov, I.D.: On a problem of Gowers. *Dokl. Akad. Nauk* **400**(2), 169–172 (2005). (Russian)
19. Shkredov, I.D.: On a problem of Gowers. *Izv. Ross. Akad. Nauk Ser. Mat.* **70**(2), 179–221 (2006). (Russian)
20. Solymosi, J.: Roth-type theorems in finite groups. *European J. Combin.* **34**(8), 1454–1458 (2013). DOI 10.1016/j.ejc.2013.05.027. URL <http://dx.doi.org/10.1016/j.ejc.2013.05.027>
21. Tao, T., Vu, V.: Additive combinatorics. Cambridge University Press, Cambridge (2006)